·特稿·

智能全科医生评估指标体系构建

赵亚利[1]，路孝琴[1]，刘珏[2]，张艺帆[3]，朱祖懿[3]，陈开元[4]，刘民[2*]，梁万年[4*]

1.100069 北京市，首都医科大学全科医学与继续教育学院
2.100191 北京市，北京大学公共卫生学院
3.100084 北京市，北京百川智能科技有限公司
4.100084 北京市，清华大学万科公共卫生与健康学院
*通信作者：梁万年，教授/博士生导师；E-mail：liangwn@tsinghua.edu.cn
刘民，教授/博士生导师；E-mail：liumin@bjmu.edu.cn
注：赵亚利与路孝琴为共同第一作者

【摘要】 **背景** 开发并推广智能医疗辅助信息系统是加强全科医生服务能力的重要手段，规范并评价智能辅助系统开发的性能和效果尤为迫切。**目的** 构建智能全科医生评估指标体系，为评估智能全科医生的服务能力提供科学工具。**方法** 2024年12月—2025年1月，通过文献研究和德尔菲专家咨询法，构建智能全科医生评估指标体系，并确定其指标权重。**结果** 三轮专家咨询的积极系数均为100%，权威系数>0.8，各级指标的重要性协调系数分别为0.210、0.255、0.145，可行性协调系数分别为0.353、0.245、0.150，最终确定的考核指标体系包括5个一级指标、11个二级指标、47个三级指标。一级指标分别为专业知识、基本医疗服务能力、预防服务能力、医患沟通与医学伦理、教育学习与科研能力，其权重分别为0.169、0.306、0.239、0.145、0.141。**结论** 本研究构建了智能全科医生评估指标，对进一步规范智能全科医生的开发及应用提供参考范围，对科技赋能提高全科医生临床实践水平具有重要现实意义。

【关键词】 人工智能；全科医生；评估指标体系；专家咨询
【中图分类号】 R-05 【文献标识码】 A DOI：10.12114/j.issn.1007-9572.2025.0020

The Construction of Assessment Index System of Artificial Intelligence General Practitioner

ZHAO Yali[1]，*LU Xiaoqin*[1]，*LIU Jue*[2]，*ZHANG Yifan*[3]，*ZHU Zuyi*[3]，*CHEN kaiyuan*[4]，*LIU Min*[2]，*LIANG Wannian*[4*]
1.School of General Medicine and Continuing Education，Capital Medical University，Shenzhen 518060，China
2.School of Public Health，Peking University，Beijing 100091，China
3.Beijing Baichuan Intelligent Technology Co.，Ltd，Beijing 100084，China
4.Vanke School of Public Health and Health，Tsinghua University，Beijing 100084，China
**Corresponding authors: LIANG Wannian，Professor/Doctoral supervisor；E-mail：liangwn@tsinghua.edu.cn*
LIU Min，Professor/Doctoral supervisor；E-mail：liumin@bjmu.edu.cn

【Abstract】 **Background** Developing and promoting intelligent medical assistance information systems is an important means to strengthen the capabilities of general practitioners. It is particularly urgent to standardize and evaluate the performance and effectiveness of intelligent assistance system development. **Objective** To construct an evaluation index system for intelligent general practitioners，and provide scientific tools for evaluating the service ability of intelligent general practitioners. **Methods** From December in 2024 to January in 2025，the evaluation index system for intelligent general practitioners with indicator weights was constructed through literature review and Delphi expert consultation. **Results** The positive coefficient of the three rounds of expert consultation was 100%，the authority coefficient was > 0.8. The coefficients of concordance for the importance of indicators at various levels were 0.210，0.255，and 0.145，respectively，while those for feasibility were 0.353，0.245，and 0.150. The

基金项目：科技创新2030——"新一代人工智能"重大项目（2021ZD0114100）；清华大学文科建设"双高"计划项目-AI赋能基层医疗服务研究（2024TSG06402）

引用本文：赵亚利，路孝琴，刘珏，等.智能全科医生评估指标体系构建[J].中国全科医学，2025. DOI：10.12114/j.issn.1007-9572.2025.0020.[Epub ahead of print][www.chinagp.net]

ZHAO Y L，LU X Q，LIU J，et al. The construction of assessment index system of artificial intelligence general practitioner[J].Chinese General Practice，2025.[Epub ahead of print]



final evaluation index system comprised five first-level indicators, 11 second-level indicators, and 47 third-level indicators. The first-level indicators were professional knowledge, basic medical service, active clinical prevention service, physician-patient communication and medical ethics, and education, learning, and research, with weights of 0.169, 0.306, 0.239, 0.145, and 0.141, respectively. **Conclusion** This study established the evaluation indicators for intelligent general practitioners, providing a reference for further standardizing the development and application of intelligent general practitioners. It has significant practical implications for enhancing the clinical practice level of general practitioners through technological empowerment..

【Key words】 Artificial Intelligence; General Practitioner; Evaluation Index System; Expert Consultation

《深化医药卫生体制改革 2024 年重点工作任务》（国办发〔2024〕29 号）强调，要加强基层医疗卫生服务能力建设，改善基层医疗卫生机构基础设施条件，推广智慧医疗辅助信息系统。全科医生作为基层医疗卫生服务的重要角色，面临着工作负担重、服务能力不均衡等问题，亟需新技术的支持来提升诊疗效率与准确性。智能全科医生（artificial intelligence general practitioner，AIGP）通过模拟人类医生的思维方式，结合大数据和机器学习技术，可以辅助全科医生进行临床决策，提高医疗服务的质量与效率[1]。基于此，为促进智能科技赋能全科医生临床实践、提升基层卫生智慧化服务水平，2024 年《智能全科医生中国专家共识》[2]提出了 17 条专家共识。在此背景下，要充分发挥 AIGP 的潜力，必须建立一套科学、全面的评估指标体系，以准确评价其性能和效果，对于推动 AIGP 的发展和应用具有重要意义。

1 资料与方法

1.1 AIGP 评估指标初步构建

通过文献复习法、政策文件梳理及专家咨询，初步形成 AIGP 评估指标体系。

（1）文献复习法：查阅分析人工智能技术、全科医生岗位胜任力等文献，提炼指标，组建指标池。

（2）政策文件梳理：检索与基层医疗卫生机构人工智能相关的政策文件，提炼相关内容，补充指标池。

（3）专家咨询法：2024 年 11 月，咨询人工智能、全科医学、基层卫生服务机构等领域专家，围绕指标池的设置及其内涵提出建议并修改指标池。初步形成 AIGP 评估指标体系。

1.2 AIGP 评估指标体系确定

在评估指标初稿基础上，展开德尔菲专家咨询法，对指标进行筛选及确定。

1.2.1 遴选专家：选取对人工智能、信息化、基层卫生服务较为熟悉的 16 位专家参与德尔菲专家咨询。专家纳入标准：（1）从事基层卫生服务、卫生管理及相关工作，且工作年限≥ 10 年；（2）具有副高级及以上职称；（3）对本研究知情且愿意参与。

1.2.2 制定专家咨询表：将初步拟定的评价指标制成电子版专家咨询问卷，包含 6 个一级指标，13 个二级指标，73 个三级指标。专家对指标的重要性和可操作性进行独立评价，评分标准为 1–5 分，1 分代表“很差”，5 分代表“很好”；同时请专家对指标的熟悉程度（Ca）和判断依据（Cs）进行自评。

1.2.3 实施专家咨询：于 2024 年 12 月—2025 年 1 月，邀请 16 名专家进行三轮德尔菲专家咨询。第一轮咨询后，对各指标重要性及可操作性的赋值均数、变异系数进行计算，将重要性、可操作性赋值均值 >3.5、变异系数≤ 0.30[3]作为筛选标准，符合两项标准之一的指标予以删除。然后结合专家具体建议对指标进行进一步修改，形成第二轮和第三轮专家咨询问卷，请专家再次打分。第二轮和第三轮指标筛选标准与第一轮相同。

1.2.4 专题小组讨论：2025 年 1 月，召开专题小组讨论，9 名专家来自人工智能、全科医学、卫生管理等领域，对德尔菲专家咨询三轮后的指标内涵进一步讨论确定。

1.3 统计学方法

采用 EpiData 3.1 和 SPSS 27.0 进行数据的录入和分析，计算各指标的均值、标准差、变异系数和肯德尔（Kendall）协调系数，其中 $P<0.05$ 为差异有统计学意义。各指标权重通过简单统计模型法确定[4]。

2 结果

2.1 咨询专家的基本情况

参与三轮咨询的专家均为 16 名，其中女性 12 名，男性 4 名；平均年龄为（50.2 ± 6.1）岁（38~63 岁）；学历以硕士及以上为主（13 人）；专业主要为临床医学（11 人）；工作年限大于 20 年的占多数（10 人）。详见表 1。

2.2 专家积极系数

本研究采用问卷回收率来表示专家的积极系数。在三轮咨询中均发放问卷 16 份，回收 16 份，有效问卷 16 份，专家积极系数为 100%。

2.3 专家权威程度

通过专家权威系数（Cr）来表示，该系数由专家判断依据系数（Ca）和专家熟悉程度系数（Cs）两个因素决定，即 $Cr=(Ca+Cs)/2$。本研究计算了一级指标、二级指标和三级指标的专家权威系数，参加专家咨询的

表1 德尔菲咨询专家基本信息情况

Table 1 The basic information of Delphi consulting experts

项目	人数	项目	人数
性别		年龄	
男	4	≤ 50 岁	8
女	12	>50 岁	8
学历		从事专业	
本科	3	临床医学（含全科医学）	11
硕士研究生	6	公共卫生	2
博士研究生	7	卫生管理	2
职称		医学伦理学	1
副高	5	工作单位	
正高	11	综合医院	1
目前岗位工作年限		社区卫生服务中心（站）	7
≤ 20 年	6	高校	7
20~30 年	7	社管中心	1
>30 年	3		

16 位专家中，5 位专家对全科医学领域信息化应用很熟悉，9 位表示比较熟悉，2 位表示一般熟悉。Cs 为 0.84，Ca 为 0.91，Cr 为 0.94。

2.4 专家意见的协调程度

通过指标评分结果的肯德尔协调系数和变异系数表示。其中，Kendall 系数取值范围为 0~1，其值越大表明协调程度越高。本研究经三轮专家咨询后，各级指标的重要性 Kendall 协调系数分别为 0.210、0.255、0.145，可行性 Kendall 协调系数分别为 0.353、0.245、0.150，其结果详见表 2。变异系数（CV）= 标准差 / 均数，CV 越小表明专家的协调程度越高，详见表 3~5。

表2 专家意见的肯德尔协调系数情况

Table 2 The Kendall coordination coefficient of the expert s' opinions

指标	重要性			可操作性		
	Kendall's *W* 值	χ^2 值	*P* 值	Kendall's *W* 值	χ^2 值	*P* 值
第一轮						
一级指标	0.138	11.000	0.049	0.322	25.729	<0.001
二级指标	0.250	39.977	<0.001	0.272	43.572	<0.001
三级指标	0.152	123.727	<0.001	0.288	235.048	<0.001
第二轮						
一级指标	0.125	10.000	0.075	0.323	25.833	<0.001
二级指标	0.144	23.111	0.010	0.233	37.298	<0.001
三级指标	0.103	87.570	0.002	0.206	175.079	<0.001
第三轮						
一级指标	0.210	13.455	0.009	0.353	22.583	<0.001
二级指标	0.255	40.807	<0.001	0.245	39.207	<0.001
三级指标	0.145	111.232	<0.001	0.150	115.558	<0.001
总体	0.170	173.873	<0.001	0.170	174.334	<0.001

2.5 AIGP 评估指标咨询结果

2.5.1 第一轮专家咨询：第一轮专家咨询表中包含 6 个一级指标，11 个二级指标，52 个三级指标。结合专家意见和指标筛选标准，对指标调整情况如下：

（1）删除指标 6 个：2–3 临终关怀服务、2–2–2 辅助家庭访视、2–3–1 对患者临终关怀服务、2–3–2 对家属情感支持、5–1–5 重视人文关怀元素、6–2–3 智能化应用的伦理责任培训；

（2）新增指标 4 个：3–1–6 重型精神病智能化健康服务、3–1–9 辅助提供重大公共卫生服务、5–1–4 服务风险识别与防范、6–2–3 持续学习和更新能力；

（3）合并指标 4 个：将 4–1–2 向患者解释病情，4–1–3 倾听患者的需求和担忧，4–1–5 向患者进行反馈、4–1–6 理解、信任并尊重患者及家属合并为 4–1–1 倾听与共情。

2.5.2 第二轮专家咨询：第二轮专家咨询表包含 6 个一级指标，10 个二级指标，48 个三级指标。经专家咨询后指标调整情况如下：

（1）合并指标 2 个：将 4 人际沟通能力、5 职业精神与人文素养合并为 4 医患沟通与医学伦理；

（2）调整指标 1 个：将 1–1–2 全科医学临床服务相关专业知识调整为 1–2 全科医学临床相关知识；

（3）新增指标新增 5 个：1–2–1 与疾病诊疗相关的各临床学科知识、1–2–2 中医和康复相关知识、1–2–3 临床服务常用技能操作规范、1–2–4 临床服务新进展及最新临床循证指南、2–2–3 家庭健康评估；

（4）删除指标 3 个：3–1–7 数智慢性病管理、3–2–5 个性化心理支持、5–1–6 医疗安全意识。

2.5.3 第三轮专家咨询：第三轮专家咨询表包含 5 个一级指标、11 个二级指标和 47 个三级指标。均通过筛选标准，且专家意见一致，均予以保留。

2.6 AIGP 评估指标体系的权重结果

根据第三轮专家咨询评分结果确定了指标体系的权重，各一级指标权重为 0.141~0.306，各二级指标权重为 0.038~0.267，各三级指标权重为 0.016~0.022，详见表 6。

3 讨论

3.1 评估指标体系构建的重要性与必要性

加强 AI 赋能基层能力建设是打通卫生健康服务“最后一公里”的重要手段[5]，尤其为边远农村地区居民提供高质量医疗服务的可及性尤为重要。随着人工智能技术的飞速发展，AIGP 在基层医疗卫生领域的应用前景广阔，将推动基层医疗卫生服务的智能化、个性化和普及化；支持临床医生完成一系列重要任务，使高质量照护更加普及[6]，成为全科医生的得力助手；同时，

表 3 一级指标的重要性、可操作性评分结果

Table 3 The importance and feasibility scores of first-level indicators

一级指标	第一轮				第二轮				第三轮			
	重要性		可操作性		重要性		可操作性		重要性		可操作性	
	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*
1 专业知识	4.94	0.05	4.81	0.11	4.94	0.05	4.81	0.11	5.00	0.00	4.94	0.05
2 基本医疗服务能力	5.00	0.00	4.94	0.05	4.94	0.05	4.75	0.12	4.94	0.05	4.56	0.16
3 预防服务能力	4.94	0.05	4.50	0.16	4.88	0.07	4.50	0.18	4.94	0.05	4.69	0.13
4 人际沟通能力	4.81	0.08	4.00	0.26	4.75	0.12	4.06	0.21	4.88	0.07	4.13	0.20
5 职业精神与人文素养	4.81	0.08	3.94	0.30	4.69	0.13	4.00	0.24	—	—	—	—
6 教育、学习与科研能力	4.69	0.13	4.44	0.16	4.63	0.13	4.44	0.20	4.69	0.10	4.44	0.14

注：*CV*= 变异系数；—表示无相关数据。

表 4 二级指标的重要性、可操作性评分结果

Table 4 The importance and feasibility scores of secondary indicators

二级指标	第一轮				第二轮				第三轮			
	重要性		可操作性		重要性		可操作性		重要性		可操作性	
	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*
1-1 全科医学基本理论、基本知识	4.94	0.05	4.81	0.08	4.94	0.05	4.81	0.11	5.00	0.00	4.94	0.05
1-2 全科医学临床相关知识	–	–	–	–	–	–	–	–	5.00	0.00	4.88	0.07
2-1 临床诊疗服务	5.00	0.00	4.81	0.08	4.94	0.05	4.69	0.13	4.94	0.05	4.63	0.11
2-2 家庭保健服务	4.75	0.09	4.13	0.21	4.88	0.07	4.38	0.20	4.81	0.08	4.44	0.14
2-3 临终关怀服务	4.56	0.11	3.19	0.38	—	—	—	—	—	—	—	—
3-1 基本及重大公共卫生服务	4.88	0.07	4.38	0.18	4.81	0.11	4.38	0.20	4.94	0.05	4.56	0.14
3-2 个性化健康管理服务	4.88	0.07	4.31	0.18	4.88	0.07	4.44	0.18	4.94	0.05	4.50	0.14
4-1 医患沟通	4.81	0.08	3.63	0.33	4.69	0.13	3.81	0.27	4.81	0.08	4.13	0.21
5-1 医学伦理与法律	4.81	0.11	3.94	0.24	4.69	0.15	4.25	0.20	4.88	0.07	4.44	0.16
6-1 辅助教学能力	4.31	0.18	3.81	0.29	4.63	0.16	4.19	0.28	4.56	0.16	4.19	0.16
6-2 自主学习能力	4.81	0.08	4.38	0.16	4.81	0.11	4.50	0.16	4.69	0.13	4.44	0.14
6-3 辅助科学研究能力	4.44	0.16	4.31	0.18	4.63	0.16	4.38	0.18	4.44	0.18	4.25	0.18

注：—表示无相关数据。

AIGP 能为全科医生提供实战培训，提升其知识和技能，是全科医生继续教育的重要工具。在推进 AIGP 的研究和实践探索过程中，需要构建一套 AIGP 评估指标体系来为 AIGP 的开发提供方向与标准，为 AIGP 的科学运用进行规范指导，对确保提升 AIGP 的安全性及合规性、确保其实际应用效果以及进一步优化 AIGP 功能具有重要意义。

3.2 评估指标体系构建过程科学、合理

本研究以政策为导向，对基层医疗卫生机构的政策文件进行了梳理，提炼出相关要素，为构建 AIGP 评估指标奠定基础，确保整个指标体系能够符合国家对基层全科医生岗位胜任力及人工智能发展的要求。同时，本研究在《智能全科医生中国专家共识》框架指引下，参考相关人工智能技术发展[7-8]及全科医生岗位胜任力评估指标相关文献[9-11]，提炼出 AIGP 应具备的能力要素，邀请来自社区卫生服务中心、高等院校、综合医院及社区卫生服务管理中心等机构对人工智能及基层卫生熟悉的专家参与指标体系构建过程，经过三轮专家咨询后最终确定 AIGP 评估指标体系。三轮专家函询问卷应答率均为 100%，专家整体积极性好，对研究的重视程度高；专家权威系数为 0.94，表明专家权威程度高，意见可靠[12]。在专家意见协调程度方面，由于第一轮及第二轮函询后，指标体系均有较大程度改动，专家意见难以迅速达成共识，导致第二轮函询中肯德尔协调系数比第一轮低，但向专家提供更多人工智能背景信息后，第三轮专家函询的肯德尔协调系数得到提升，最终三轮专家意见协调系数的 P 值均 <0.05，说明各指标间的协调程度较好[12]。

3.3 指标体系全面考核五大核心能力

本评估指标体系对 AIGP 的五大能力进行考核，包括专业知识、基本医疗服务能力、预防服务能力、医患沟通与医学伦理，以及教育、学习与科研能力。

表 5 三级指标的重要性、可操作性评分结果

Table 5 The importance and feasibility scores of tertiary indicators

三级指标	第一轮				第二轮				第三轮			
	重要性		可操作性		重要性		可操作性		重要性		可操作性	
	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*
1–1–1 全科医学基本概念、基本理论和原则	4.69	0.13	4.63	0.16	4.94	0.05	4.63	0.13	4.88	0.07	4.63	0.11
1–1–2 全科医学相关人文社会科学知识	4.94	0.05	4.75	0.12	4.94	0.05	4.75	0.12	–	–	–	–
1–1–3 全科医学相关预防医学知识	4.81	0.11	4.56	0.16	4.88	0.07	4.50	0.16	4.88	0.07	4.63	0.11
1–1–4 医学相关科学研究知识	4.88	0.07	4.81	0.11	4.88	0.07	4.81	0.11	4.75	0.09	4.56	0.14
1–2–1 与疾病诊疗相关的各临床学科知识	—	—	—	—	—	—	—	—	4.88	0.07	4.63	0.13
1–2–2 中医和康复相关知识	—	—	—	—	—	—	—	—	4.75	0.09	4.25	0.18
1–2–3 临床服务常用技能操作规范	—	—	—	—	—	—	—	—	4.81	0.08	4.13	0.20
1–2–4 临床服务新进展及最新临床循证指南	—	—	—	—	—	—	—	—	4.88	0.07	4.69	0.13
2–1–1 智能辅助进行病史采集	5.00	0.00	4.69	0.10	4.94	0.05	4.69	0.13	4.81	0.08	4.63	0.13
2–1–2 智能获取或辅助体格检查	5.00	0.00	4.38	0.16	4.94	0.05	4.50	0.14	4.75	0.12	4.13	0.17
2–1–3 智能辅助检查的选择与结果解读	5.00	0.00	4.63	0.13	4.88	0.07	4.63	0.16	4.88	0.07	4.56	0.11
2–1–4 智能辅助诊断和鉴别诊断	4.94	0.05	4.63	0.11	4.88	0.07	4.63	0.11	4.88	0.07	4.56	0.11
2–1–5 健康咨询	4.88	0.07	4.38	0.16	4.81	0.08	4.44	0.16	4.81	0.08	4.44	0.14
2–1–6 智能辅助制定治疗方案	4.88	0.07	4.56	0.14	4.88	0.07	4.69	0.10	4.81	0.08	4.38	0.16
2–1–7 智能辅助进行危重症处理	4.81	0.08	3.56	0.32	4.88	0.07	4.31	0.25	4.81	0.08	4.38	0.14
2–1–8 智能辅助合理用药	4.94	0.05	4.56	0.16	4.94	0.05	4.63	0.13	4.81	0.08	4.50	0.11
2–1–9 智能辅助个性化康复及照护服务	4.75	0.09	3.94	0.24	4.88	0.07	4.25	0.20	4.75	0.09	4.25	0.14
2–1–10 智能辅助精准转诊	4.88	0.07	4.19	0.25	4.88	0.07	4.63	0.16	4.75	0.12	4.38	0.16
2–1–11 远程医疗照护服务	4.88	0.07	4.31	0.24	4.69	0.13	4.38	0.23	4.50	0.14	4.19	0.20
2–1–12 随访及慢性病监测服务	4.88	0.07	4.31	0.18	4.88	0.07	4.50	0.18	4.63	0.13	4.50	0.16
2–1–13 辅助医疗文书撰写	4.88	0.07	4.63	0.11	4.88	0.07	4.56	0.16	4.63	0.11	4.44	0.16
2–2–1 家庭健康咨询	4.81	0.08	4.25	0.20	4.81	0.08	4.69	0.13	4.63	0.11	4.50	0.11
2–2–2 辅助家庭访视	4.31	0.26	3.63	0.32	—	—	—	—	—	—	—	—
2–2–3 家庭健康评估	—	—	—	—	—	—	—	—	4.56	0.14	4.38	0.16
2–3–1 对患者临终关怀服务	4.38	0.18	3.13	0.35	—	—	—	—	—	—	—	—
2–3–2 对家属情感支持	4.31	0.14	3.06	0.33	—	—	—	—	—	—	—	—
3–1–1 辅助居民健康档案建立和使用	4.94	0.05	4.81	0.08	4.88	0.07	4.69	0.13	4.81	0.08	4.69	0.13
3–1–2 健康宣教	4.56	0.14	3.81	0.29	4.38	0.25	4.13	0.26	4.56	0.14	4.56	0.14
3–1–3 孕产妇智能化保健服务	4.63	0.11	4.13	0.20	4.56	0.16	4.19	0.22	4.63	0.16	4.25	0.18
3–1–4 0~6 岁儿童智能化保健服务	4.69	0.10	4.00	0.20	4.50	0.18	4.13	0.23	4.75	0.12	4.44	0.16
3–1–5 老年人智能化健康服务	4.75	0.09	4.00	0.26	4.88	0.07	4.56	0.14	4.81	0.08	4.44	0.14
3–1–6 重型精神病智能化健康服务	—	—	—	—	4.56	0.18	4.13	0.28	4.31	0.20	3.81	0.24
3–1–7 数智慢性病管理	4.94	0.05	4.31	0.14	4.88	0.07	4.69	0.10	—	—	—	—
3–1–8 辅助传染病及突发公共卫生事件报告和处置	4.63	0.13	3.69	0.27	4.75	0.09	4.38	0.16	4.56	0.16	4.31	0.18
3–1–9 辅助提供重大公共卫生服务	—	—	—	—	4.56	0.18	4.25	0.22	4.50	0.16	4.31	0.22
3–2–1 个性化保健服务包的制定	4.81	0.11	4.25	0.16	4.75	0.09	4.38	0.14	4.63	0.13	4.44	0.16
3–2–2 健康风险因素评估及预测	4.88	0.07	4.81	0.11	4.81	0.08	4.56	0.14	4.75	0.12	4.63	0.16
3–2–3 中医药健康管理服务	4.56	0.14	3.63	0.28	4.75	0.09	4.13	0.20	4.50	0.16	4.31	0.20
3–2–4 体检规划	4.81	0.08	4.75	0.12	4.75	0.12	4.56	0.16	4.75	0.09	4.69	0.10
3–2–5 个性化心理支持	4.75	0.12	3.63	0.36	4.81	0.11	3.75	0.33	—	—	—	—
4–1–1 倾听与共情	4.75	0.12	3.81	0.32	4.88	0.07	3.56	0.34	4.63	0.11	4.00	0.18
4–1–2 向患者解释病情	4.88	0.07	3.69	0.31	—	—	—	—	—	—	—	—
4–1–3 倾听患者的需求和担忧	4.69	0.13	3.44	0.37	—	—	—	—	—	—	—	—
4–1–4 辅助医患共同决策	4.81	0.08	4.19	0.20	4.88	0.07	3.88	0.32	4.69	0.13	4.25	0.16
4–1–5 向患者进行反馈	4.81	0.08	4.25	0.20	—	—	—	—	—	—	—	—
4–1–6 理解、信任并尊重患者及家属	4.50	0.16	3.00	0.39	—	—	—	—	—	—	—	—



三级指标	第一轮				第二轮				第三轮			
	重要性		可操作性		重要性		可操作性		重要性		可操作性	
	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*	$\bar{x}$	*CV*
5-1-1 遵守医疗法律法规	4.81	0.11	4.50	0.16	4.94	0.05	4.69	0.13	4.88	0.07	4.69	0.13
5-1-2 遵守职业道德	4.88	0.07	4.25	0.22	4.88	0.07	4.50	0.18	4.88	0.07	4.56	0.16
5-1-3 维护患者知情权和隐私权	4.81	0.08	4.19	0.23	4.88	0.07	4.44	0.25	4.75	0.12	4.38	0.20
5-1-4 服务风险识别与防范	—	—	—	—	4.88	0.07	4.75	0.12	4.81	0.08	4.50	0.16
5-1-5 重视人文关怀元素	4.75	0.12	3.50	0.35	—	—	—	—	—	—	—	—
5-1-6 医疗安全意识	4.88	0.07	4.06	0.26	4.94	0.05	4.44	0.20	—	—	—	—
6-1-1 辅助示教能力	4.56	0.20	4.00	0.33	4.81	0.11	4.44	0.20	4.69	0.10	4.56	0.14
6-1-2 模拟实践培训	4.56	0.20	4.25	0.30	4.88	0.07	4.44	0.18	4.63	0.11	4.63	0.13
6-2-1 个性化课程学习	4.69	0.17	4.38	0.26	4.88	0.07	4.38	0.20	4.63	0.13	4.50	0.16
6-2-2 智能化应用的伦理责任培训	4.44	0.22	3.81	0.37	—	—	—	—	—	—	—	—
6-2-3 持续学习和更新能力	—	—	—	—	4.88	0.07	4.56	0.18	4.69	0.13	4.31	0.18
7-3-1 辅助检索文献、解读和记录信息的能力	4.88	0.07	4.81	0.11	4.88	0.07	4.69	0.15	4.50	0.16	4.63	0.16
7-3-2 辅助撰写科研报告	4.75	0.09	4.69	0.13	4.75	0.16	4.56	0.18	4.44	0.16	4.38	0.18

注：—表示无相关数据。

首先，一级指标中“基本医疗服务能力”的权重占比最高，是 AIGP 最核心的能力，其二三级指标全面反映 AIGP 在实际临床应用中应涵盖的重要服务项目，协助全科医生提供综合、优质的基本医疗服务，是衡量 AIGP 创新能力与竞争优势的关键指标。同时将服务对象从个体拓展到家庭，体现出全科医生连续性服务的特色。现阶段，AIGP 的服务锚点应满足社区主要需求，以“智能辅助进行病史采集”“智能辅助检查选择与结果解读”“智能辅助诊断和鉴别诊断”几方面为主进行精进，同时随着 AIGP 发展的不断更新迭代，进一步扩展涵盖其他基本医疗服务项目。其次，一级指标中权重占比第二位的是“预防服务能力”。2024 年 7 月 18 日，党的二十届三中全会审议通过《中共中央关于进一步全面深化改革、推进中国式现代化的决定》，提出“促进社会共治、医防协同、医防融合”。基层医疗卫生机构作为基本医疗和基本公共卫生服务的网底，是实现“医防融合”的主要场所之一，而全科医生是实现基层“医防融合”的最佳人选。利用 AIGP 的巨大优势，辅助全科医生在日常诊疗工作中，及时主动为社区居民提供预防服务，助力全科医生提升“医防融合”服务能力，促进“医防融合”在社区的有效落地，是 AIGP 在发展和应用过程中的必备要素。在三级指标中，“辅助居民健康档案建立和使用”及“体检规划”的权重最高，被认为是现阶段 AIGP 在减轻全科医生管理工作负担中最需要实现的重要任务。

具备全面且专业的全科相关知识，是 AIGP 体现全科理念，进而有效辅助全科医生、实现良好应用效果的重要保障。强大的人工智能技术能够支撑 AIGP 不断通过自我学习、自我校正来更新专业知识、完善模块功能以适应不同基层场景的变化发展。

在医患沟通及医学伦理指标上，需要通过设定医学伦理与法律法规要求，以保证 AIGP 使用过程中的安全性与合规性[13]。同时以需求为导向实现良好的医患沟通，不断提升 AIGP 易用性及患者体验感[14]。但是从三级指标“倾听与共情”“辅助医患共同决策”权重分值及可操作性评分结果来看，AIGP 在医患沟通方面的应用效果及优化需要时间来验证。

最后的教学与科研能力指标，旨在通过评价 AIGP 在医学教育与培训中的效果，来推动 AIGP 进一步辅助医学教育[15]、在促进全科医生教学科研能力中发挥积极作用。

4 小结

本研究采用文献研究和德尔菲专家咨询法构建了我国首个智能全科医生评估指标体系，该指标体系共包含 5 个一级指标、11 个二级指标、47 个三级指标，为智能全科医生服务技术的评估提供了科学工具，以促进 AI 技术在全科医学领域的进一步研究及应用。本研究尚存在一定局限性，由于 AIGP 发展尚处于实践探索阶段，所构建的 AIGP 评估指标体系未来有待于进一步验证，并进行适应性调整。

作者贡献：梁万年、刘民负责课题构思、设计、实施、文章撰写等全程审核和把关；路孝琴、赵亚利、刘珏参与课题实施、文章撰写及修改；张艺帆、朱祖懿、陈开元参与课题实施。

本文无利益冲突。

参考文献

表 6　智能全科医生评估指标体系及其权重

Table 6　The evaluation index system of intelligent general practitioner and its weight

一级指标	权重	二级指标	权重	三级指标	权重
1 专业知识	0.169	1–1 全科医学基本理论、基本知识	0.085	1–1–1 全科医学基本概念、基本理论和原则	0.022
				1–1–2 全科医学相关人文社会科学知识	0.020
				1–1–3 全科医学相关预防医学知识	0.022
				1–1–4 全科医学相关科学研究知识	0.021
		1–2 全科医学临床相关知识	0.084	1–2–1 与疾病诊疗相关的各临床学科知识	0.022
				1–2–2 中医和康复相关知识	0.020
				1–2–3 临床服务常用技能操作规范	0.020
				1–2–4 临床服务新进展及最新临床循证指南	0.022
2 基本医疗服务能力	0.306	2–1 临床诊疗服务	0.267	2–1–1 智能辅助进行病史采集	0.022
				2–1–2 智能获取或辅助体格检查	0.019
				2–1–3 智能辅助检查的选择与结果解读	0.022
				2–1–4 智能辅助诊断和鉴别诊断	0.022
				2–1–5 健康咨询	0.021
				2–1–6 智能辅助制定治疗方案	0.021
				2–1–7 智能辅助进行危重症处理	0.021
				2–1–8 智能辅助合理用药	0.021
				2–1–9 智能辅助个性化康复及照护服务	0.020
				2–1–10 智能辅助精准转诊	0.020
				2–1–11 远程医疗照护服务	0.018
				2–1–12 随访及慢性病监测服务	0.020
				2–1–13 辅助医疗文书撰写	0.020
		2–2 家庭保健服务	0.039	2–2–1 家庭健康咨询	0.020
				2–2–2 家庭健康评估	0.019
3 预防服务能力	0.239	3–1 基本及重大公共卫生服务	0.157	3–1–1 辅助居民健康档案建立和使用	0.022
				3–1–2 健康宣教	0.020
				3–1–3 孕产妇智能化保健服务	0.019
				3–1–4 0~6 岁儿童智能化保健服务	0.021
				3–1–5 老年人智能化健康服务	0.021
				3–1–6 重型精神病智能化健康服务	0.016
				3–1–7 辅助传染病及突发公共卫生事件报告和处置	0.019
				3–1–8 辅助提供重大公共卫生服务	0.019
		3–2 个性化健康管理服务	0.082	3–2–1 个性化保健服务包的制定	0.020
				3–2–2 健康风险因素评估及预测	0.021
				3–2–3 中医药健康管理服务	0.019
				3–2–4 体检规划	0.022
4 医患沟通与医学伦理	0.145	4–1 医患沟通	0.038	4–1–1 倾听与共情	0.019
				4–1–2 辅助医患共同决策	0.019
		4–2 医学伦理与法律	0.107	4–2–1 遵守医疗法律法规	0.022
				4–2–2 遵守职业道德	0.022
				4–2–3 维护患者知情权和隐私权	0.020
				4–2–4 服务风险识别与防范	0.022
5 教育、学习与科研能力	0.141	5–1 辅助教学能力	0.042	5–1–1 辅助示教能力	0.021
				5–1–2 模拟实践培训	0.021
		5–2 自主学习能力	0.040	5–2–1 个性化课程学习	0.020
				5–2–2 持续学习和更新能力	0.020
		5–3 辅助科学研究能力	0.059	5–3–1 辅助检索文献、解读和记录信息能力	0.020
				5–3–2 辅助撰写科研报告	0.019



[1] 闫温馨，胡健，曾华堂，等. 人工智能大语言模型在基层医疗卫生服务中的应用与挑战[J]. 中国全科医学，2025，28 (1): 1-6.

[2] 清华大学万科公共卫生与健康学院，北京大学公共卫生学院，中国医师协会全科医师分会. 智能全科医生中国专家共识[J]. 中国全科医学，2025，28(2): 135-142.

[3] DIAMOND I R，GRANT R C，FELDMAN B M，et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies[J].J Clin Epidemiol，2014，67(4): 401-409. DOI:10.1016/j.jclinepi.2013.12.002.

[4] RAYMOND M R. Job analysis and the specification of content for licensure and certification examinations[J]. Appl Meas Educ，2001，14(4): 369-415. DOI:10.1207/s15324818ame1404_4.

[5] 刘珏，梁万年. 科技创新应对健康挑战[J]. 中国全科医学，2024,27(28): 前插页.

[6] MOOR M，BANERJEE O，ABAD Z S H，et al. Foundation models for generalist medical artificial intelligence[J]. Nature, 2023，616：259-265.

[7] LI Y H，LI Y L，WEI M Y，et al. Innovation and challenges of artificial intelligence technology in personalized healthcare[J]. Sci Rep, 2024，14(1): 18994. DOI:10.1038/s41598-024-70073-7.

[8] ANDREW A. Potential applications and implications of large language models in primary care[J]. Fam Med Community Health, 2024，12(Suppl 1): e002602. DOI:10.1136/fmch-2023-002602.

[9] 李婷，潘朝路，金光辉，等. 基于岗位胜任力的全科住院医师规范化培训结业考核指标体系的构建[J]. 中国全科医学，2024，27(16):1962-1970.

[10] 宫雪. 中国全科医生岗位胜任力指标体系构建研究[D]. 中国医科大学，2022.

[11] 连国华，陈亮，张向杰，等. 基于洋葱模型的社区全科医生岗位胜任力评价指标体系构建研究[J]. 中国全科医学，2022，25(31):3955-3959.

[12] 王媛媛. 医学量表的编制与评价：理论、方法与实例操作[M]. 北京：北京大学医学出版社，2020: 66-68.

[13] ANDREW A. Potential applications and implications of large language models in primary care[J]. Fam Med Community Health，2024，12(Suppl 1): e002602. DOI:10.1136/fmch-2023-002602.

[14] LI Y H，LI Y L，WEI M Y，et al. Innovation and challenges of artificial intelligence technology in personalized healthcare[J]. Sci Rep, 2024，14: 18994. DOI:10.1038/s41598-024-70073-7.

[15] 江哲涵，奉世聪，王维民. 人工智能生成内容在医学教育中的应用、挑战与展望[J]. 中国教育信息化，2024，30(8): 29-40.

（收稿日期：2025-02-06；修回日期：2025-03-13）
（本文编辑：李婷婷）